

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

## 4,800

Open access books available

## 122,000

International authors and editors

## 135M

Downloads

Our authors are among the

## 154

Countries delivered to

## TOP 1%

most cited scientists

## 12.2%

Contributors from top 500 universities

**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Speech Interface Evaluation on Car Navigation System – Many Undesirable Utterances and Severe Noisy Speech –

Nobuo Hataoka<sup>1</sup>, Yasunari Obuchi<sup>2</sup>,  
Teppei Nakano<sup>3</sup> and Tetsunori Kobayashi<sup>3</sup>

<sup>1</sup>*Department of Electronics and Intelligent Systems,  
Tohoku Institute of Technology,*

<sup>2</sup>*Central Research Laboratory, Hitachi Ltd.,*

<sup>3</sup>*Waseda University,  
Japan*

## 1. Introduction

Recently, ASR (Automatic Speech Recognition) functions have commercially been used for various consumer applications including car navigation systems. However, many technical and usability problems still exist before ASR applications are on real business use. Our goal is to make ASR technologies for a real business use. To do so, we first evaluate a car navigation interface which has ASR as an input method, and second evaluate an ASR module using real noisy in-car speech. For ASR applications, we envision mobile environments, e.g. mobile information service systems such as car navigation systems and cellular phones on which an embedded speech recognizer (Kokubo et. al., 2006) is running and which are connected to remote servers that support various information-seeking tasks.

Taking a look at commercially available car navigation systems, currently over 75% systems have ASR interfaces, however, there are very few drivers who have experiences to use the ASR interfaces. What is the problem? This is caused by the ASR usability problems.

In this chapter, we report two experimental evaluation results of ASR interface for mobile use, especially for car navigation applications. First, we evaluate the usability aspects of speech interface and second, we evaluate in-car noise speech problems to propose an effective method to cope with noisy speech. For the first evaluation, we use a prototype which has a promising speech interface called FlexibleShortcuts and Select&Voice produced by Waseda University (Nakano et. al., 2007). We found many undesirable OOV (Out-Of-Vocabulary) utterances which make the interface worse. From the second experiment to check car-noise problems, we propose an array microphone + Spectrum Subtraction (SS) technique to increase recognition accuracy.

## 2. Problems of ASR car use

### 2.1 System concept of network applications

As information technology expands into the mobile environment to provide ubiquitous communication, intelligent interfaces will be a key element to enable mobile access to

networked information. For mobile information access, HMIs (Human Machine Interfaces) using speech might be the most important and essential as speech interfaces are more effective for small, portable devices. Mobile terminals such as cellular phones, PDAs (Personal Digital Assistants) and Hand-held PCs (Personal Computers) are connected to networks like the Internet to access information from web servers. For mobile information access, speech processing and image processing will be key technologies on intelligent mobile terminals.

Especially, Car Telematics refers to a new service concept where mobile terminals (e.g. car navigation systems, cellular phones) are used to connect to networked information services. Figure 1 illustrates a total service system concept, which consists of three parts, e.g. terminal/client, network, and center/server. For the terminals, sophisticated HMIs are required to handle various inquiries and to deliver information from the center using speech and image input/output.

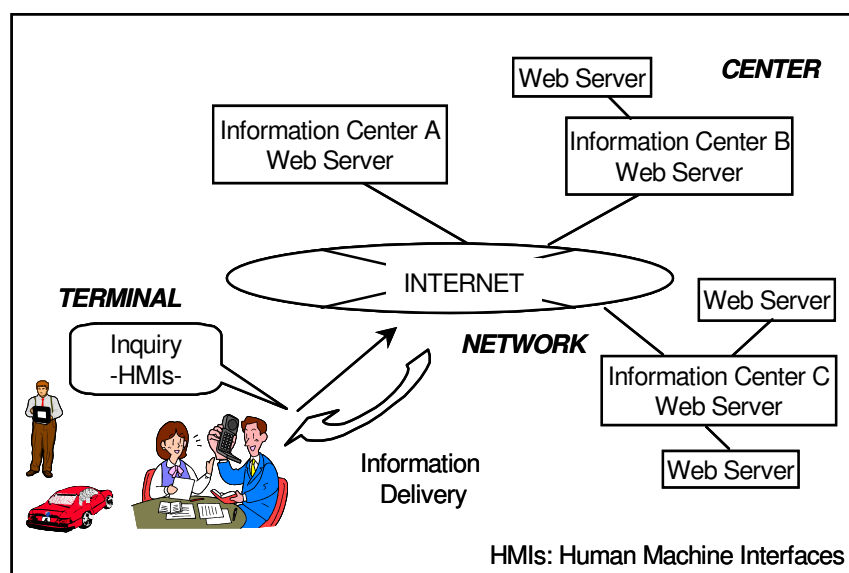


Fig. 1. Service System Image (Terminal, Network, and Center)

The network is typically the Internet; and via the Internet, the user's requests are transferred to related Web servers at the center, and required information will be provided from the center to users via networks and terminals (Hataoka et. al., 2004).

## 2.2 Technical problems for automotive use

In cars, HMIs based on speech processing such as ASR and TTS are essential to provide a safe driving environment. However, there are many problems to be solved before a real use of ASR and TTS as follows;

1. **Usability problem:** All interfaces should have a transparent navigation model. However the interfaces using ASR do not have this function/feature. If the input is misrecognized, the user can not understand why the misrecognition occurred and then can not manage the next action.
2. **OOV (out of vocabulary) problem:** There are many ways to express one meaning/location. For example, we can say either "Starbucks" or "Staba" to show the same coffee shop. It is essential for the ASR to handle this OOV problem (Vertanen, 2008)

3. **Robustness issue:** The in-car speech has noise including an engine noise and audio noise etc. To enhance the degraded speech is essential for ASR. The array microphones are used to locate sound source and reduce environmental noise (Obuchi & Hataoka, 2006).

In this chapter, to deal with these problems, first we evaluated usability issues of ASR interface on a car navigation system, and second evaluated robustness of ASR module on the car navigation interface.

### 3. ASR interface evaluation

#### 3.1 Pre-evaluation using commercial product

##### 3.1.1 Evaluation setup and task

To make real problems of ASR interfaces clear, first, we evaluate a speech interface of a commercial product (Figure 2: PIONEER Carrozzeria AVIC-HRZ88II) using two environments. The first environment is in a laboratory room and the second one is in a noisy driving car. The number of subjects is 20 people. All are university students and no one has an experience of using speech interfaces. At the beginning, each subject was instructed how to use a car navigation ASR interface by an operator.



Fig. 2. Evaluated Car Navigation System (PIONEER)

Two tasks are evaluated under two environments of the room and the in-car. The evaluation tasks are as follows;

1. Command input: audio (radio/DVD etc) and air conditioner operation, i.e. "FM radio channel 4" etc.
2. Destination setting for navigation: Two utterance ways were evaluated, first utterance from the written vocabulary and second prompt utterance (free word)

##### 3.1.2 Evaluation results

1. Room environment: Figure 3 shows evaluation results. The recognition success turn times are shown.
2. In-car environment: Figure 4 shows evaluation results of driving car environment. The results became worse than those of in room environment.

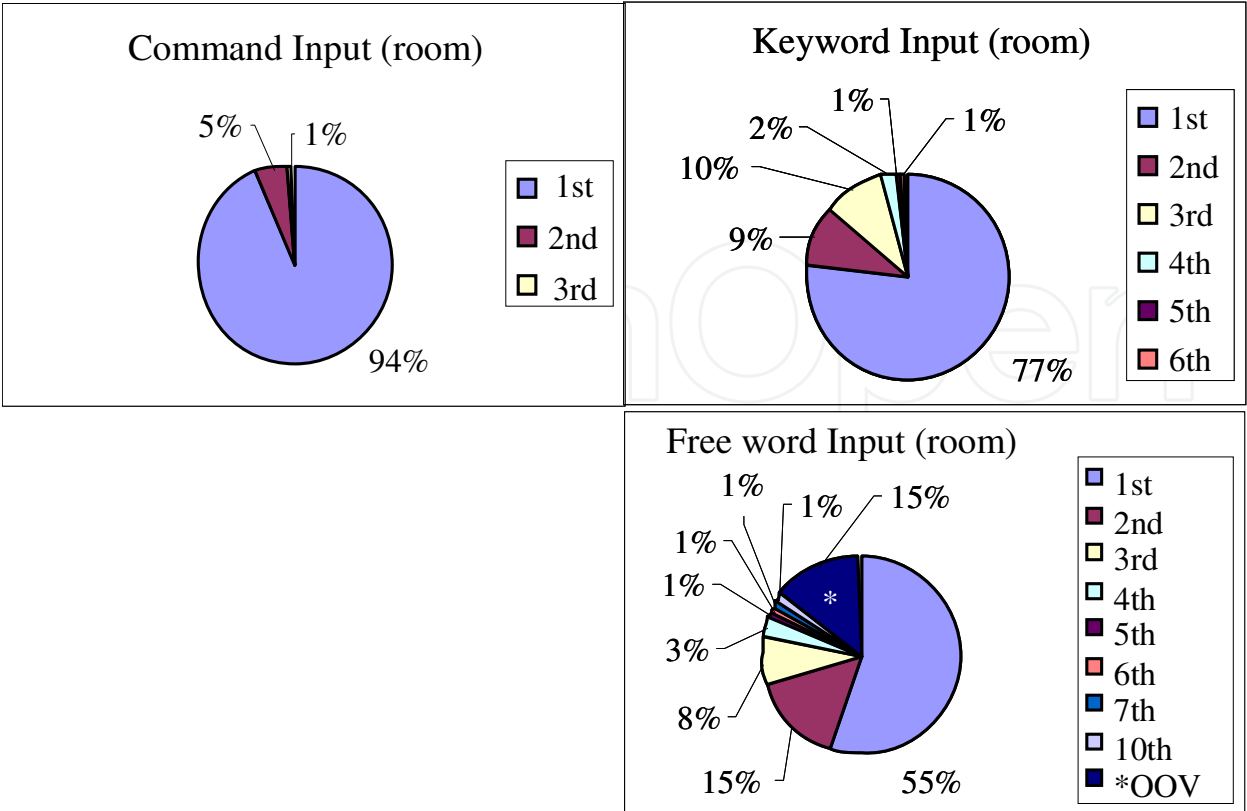


Fig. 3. Evaluation Results (laboratory room)

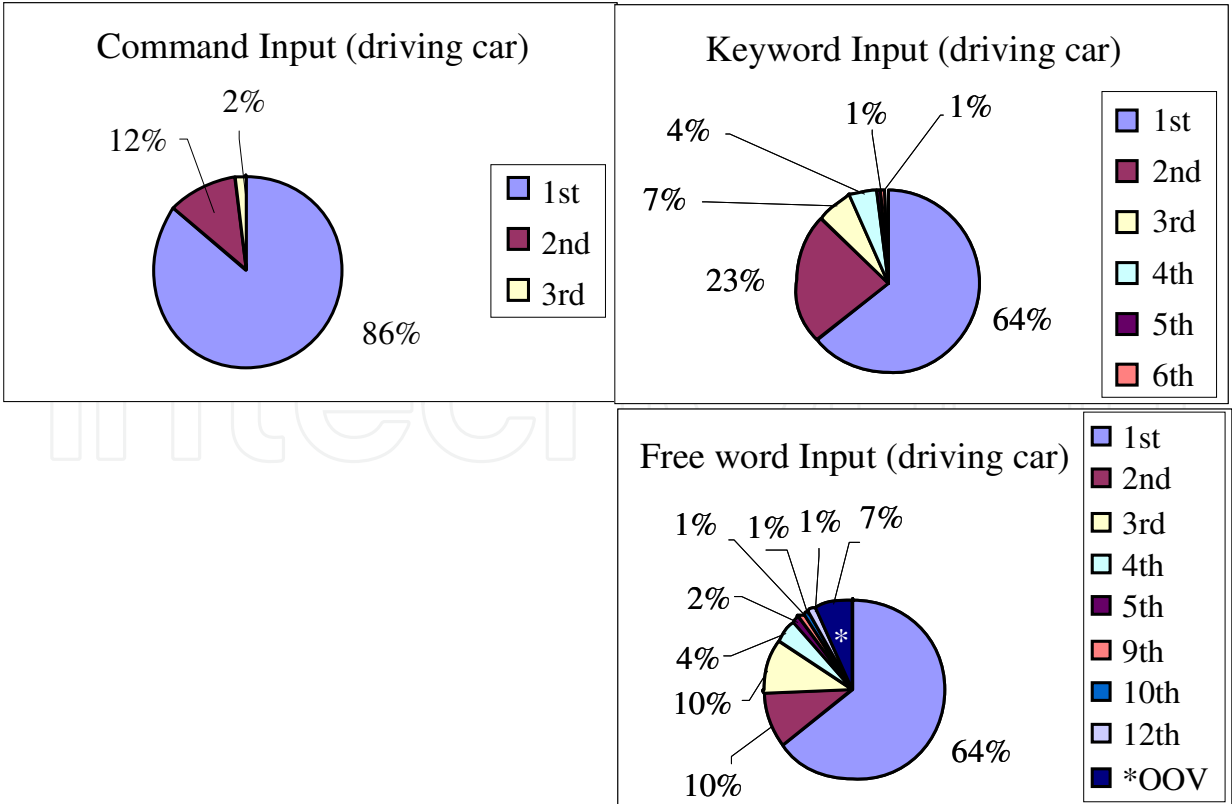


Fig. 4. Evaluation Results (driving car)

### 3.1.3 Consideration

The OOV issue was crucial which occurred at the free word utterance. Then we are developing an overcoming system to cope with this OOV problem. The system consists of terminals and centers, and when recognition errors and interruption of the speech input occur, the system sends all interaction records and speech data files to the center. At the center, a full specification of continuous ASR can recognize the data and then deliver a new vocabulary set to terminals.

## 3.2 Evaluation using prototype system

### 3.2.1 Philosophy for evaluation

First, we use the ASR prototype system called FlexibleShortcuts and Select&Voice. The FlexibleShortcuts can handle both of voice input and menu input and for the voice input many short-cuts are available to say related words directly. The Select&Voice has framed-based input windows to utter input words. The FlexibleShortcuts and Select&Voice are useful for car navigation task. Second, we use a location retrieval task in that many OOV utterances would be observed frequently in order to check how users act when the OOV utterances occur.

### 3.2.2 FlexibleShortcuts and Select&Voice

For the evaluation of OOV problems, we used a system consisting of FlexibleShortcuts and Select&Voice which have been developed by Waseda University (Nakano et. al., 2007). Waseda University is developing the Proxy-Agent as the platform of ASR application systems by the Japanese National Found Research Project called “Fundamental Technology Development on Speech Recognition.” This found was for three years from fiscal year of 2006 to 2008. The Proxy-Agent has characteristic features of plug-in based function merging and connection to network servers. The function merging is independent from ASR engines to do data collection, new ASR engine adding, and co-use of possible ASR parts etc.

The FlexibleShortcuts and the Select&Voice are the speech interface functions which are developed as an application development tool in the Proxy-Agent framework. The FlexibleShortcuts is a speech interface having flexible selection of speech inputs and/or menu inputs and also shortcut functions. In the menu expression, if a user knows the shortcut way meaning the final word to say, the user can utter this final word to reach the destination. For example, in the menu expression FM radio is under a radio category, but we can say “FM radio” directly to reach the FM radio handling process. The Select&Voice is a speech interface for data input which has been developed according to the analogy of GUI (Graphical User Interface). The Select&Voice has the framed-based input windows. This is named because the input processes are first “Select” input frame “And” then utter “Voice.”

The car navigation system which has the speech interface based on FlexibleShortcuts and Select&Voice is evaluated on the location/address retrieval task.

### 3.2.3 Evaluation experiment details

#### 1. Evaluation Purpose

In this experiment, we evaluate whether the subjects can realize their OOV utterances and if so, when they can realize OOV utterances and how they act after realization using the framework of FlexibleShortcuts+ Select&Voice. We carefully select a task in which OOV will occur easily. From this viewpoint, we select a city and town setting task in which many city



and town name changes occur after city and town merging by the local government. All actions are recorded as input and output logs using the Proxy-Agent architecture. Finally we analyze subjects' behaviors using recognition results and recorded logs.

2. Data Collection Environment

To focus on the OOV utterance and subject's action (behavior), the quiet environment of a laboratory room is set for the evaluation experiments.

3. Evaluated System/Equipment

We use the PC system consisting of FlexibleShortcuts+ Select&Voice developed by Waseda University on the car navigation task. Figure 5 shows a PC and a controller. Both menu selection and voice utterance are possible using the controller.



Fig. 5. PC and Controller used for Evaluation

4. Experimental Subjects

The number of subjects is 10 (ten) consisting of 5 (five) subjects who have experiences to use this kind of speech interface and 5 (five) subjects who have not experiences. The input and output logs and utterances are recorded using PC and a video recorder.

5. Utterance Conditions

Each subject retrieves ten (10) locations/addresses using speech utterance. In the ten locations, two names are changed by the new city merging. This means former names are OOVs. Figure 6 shows a display example of the Select&Voice interface showing prefecture, city, area, and address numbers.

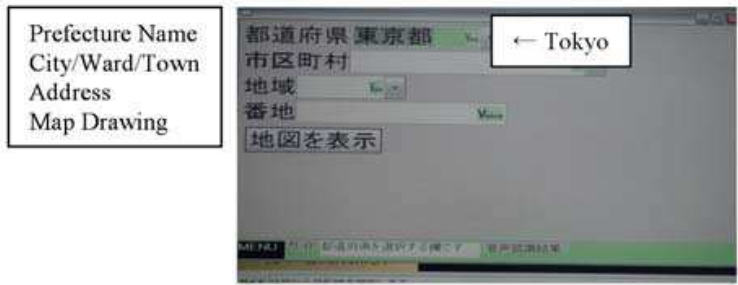


Fig. 6. Display of Location Retrieval Task

3.2.4 Evaluation experiment setup

At first, an experimental operator introduces how to use the system to subjects and then subjects start evaluation experiments after short use of the system. In the evaluation stage, there is no advice and suggestion from the operator. The display of PC is recorded using a video recorder and all utterances and input and output logs are recorded in the PC memories.

3.2.5 Evaluation results

1. Evaluation Data for Location Retrieval  
In the ten locations, two names of locations/addresses have been changed by the city merging. The following data show two names changed.

(1)before: Aijyo 1-1, Osato-machi, Osato-gun, Saitama  
changed to Aijyo 1-1, Kumagaya-shi, Saitama  
(2)before: Okisu 1-1, Kamisu-machi, Kashima-gun, Ibaragi  
changed to Okisu 1-1, Kamisu-shi, Ibaragi

Others are three ordinance-designated city names which have area’s name “ku.” We normally do not understand how to separate “ku” and the following location names. Other 5 locations/addresses are normal names.  
2. OOV Ratio

Table 1 shows evaluation results showing ratios of correct recognition, OOV, and misrecognition, respectively. In the experienced group, the total number of utterances is 352 and the number of correct recognition is 191 (54%), the number of OOV is 86 (24%), and the number of misrecognition is 74 (21%). In the non-experienced group, the total number of utterances is 601 and the numbers of correct recognition, OOV, and misrecognition are 304 (50%), 151 (26%) and 146 (24%), respectively.

Experienced			Non experienced		
	No.	Ratio	No.	Ratio	
Correct recognition	191	54%	304	50%	
OOV	86	24%	151	26%	
Mis-recognition	74	21%	146	24%	
bug	1		1		
total	352		601		

Table 1. Evaluation Results according to Experience

3. Feature of OOT Utterances  
The varieties of OOV of the location/address retrieval task are complete OOV utterances, mis-division of ordinance-designated city names, input frame errors, and OOV utterances at the top display level of the FlexibleShortcuts stage. Table 2 shows details of OOV utterances. For the total 951 utterances, the number of OOV is 237 consisting of 112 complete OOVs, 55 address division errors, 21 input frame errors, and 49 top-display-level OOVs.



Complete OOVs	Address division errors	Input frame errors	Top display level	Total
112	55	21	49	237
47%	23%	9%	21%	

Table 2. The Number of OOV Utterances

4. Subject Action/Behavior after OOV Utterance

The following remarks are subjects’ actions after OOV utterances.

a. Complete OOV utterances:

Table 3 shows an example in which a subject realized OOV after four (4) utterances (repeated). This subject realized an OOV utterance checking a vocabulary list shown using the controller. Figure 7 shows the number of utterances when subjects realized OOV utterances. The numbers of cases in which subjects realized OOVs at the second utterance, third one, forth one, fifth one, sixth one, seventh one, eighth one, and ninth one are 1 (5%), 3 (15%), 4 (20%), 1 (5%), 2 (10%), 6 (30%), 0 (0%), and 3 (15%), respectively.

Utterances	Recognition results	
Content	Recognized word	Reason
Ibaragi	Ibaragi	(correct)
Kashima-gun Kamisu-cho	Kusumigaura-shi	OOV
Kashima-gun Kamisu-cho	Sarushima-gun, Gosumi-cho	OOV
Kashima-gun Kamisu-cho	Sarushima-gun, Sakae-cho	OOV
Kashima-gun Kamisu-cho	Sarushima-gun, Sakae-cho	OOV
To the next address retrieval step		

Table 3. Example of OOV Utterances (address task)

b. Address Division Errors

To retrieve ordinance-designated cities, users should utter city name and “ku” name together, however most people utter city name and “ku” name separately. (Example: Sapporo-shi and Chuou-ku separately instead of Sapporo-shi Chuou-ku) Firstly, subjects utter these city and “ku” names separately, but when connected error output of city and “ku” names appeared on the display subjects understood how to utter ordinance-designated city names. Finally subjects could retrieve locations/addresses correctly. This is advantage of the Select&Voice architecture based on input frames.

c. Input Frame Errors

Subjects should select input frames correctly. If not so, subject’s utterance will be an OOV utterance. From the check of logs, subjects can change the input frames correctly after realizing the misrecognition results in this case.

The improvement of ASR interfaces and how to deal with OOV utterances are the most important and urgent issues to be solved. To create new vocabularies (the same meaning, but different utterance styles) from original vocabulary automatically is one of future research issues.

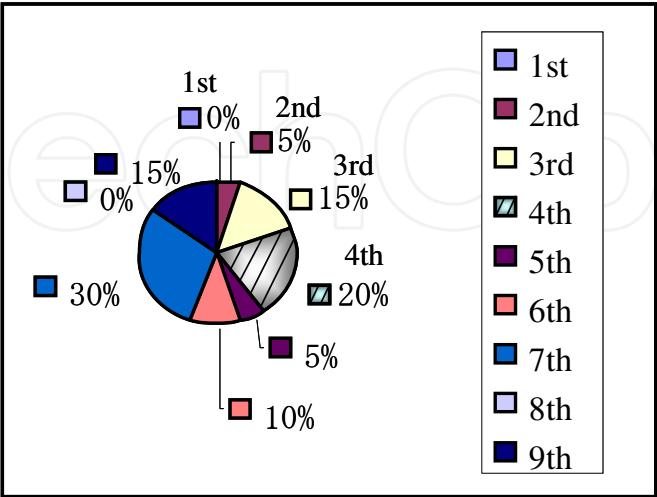


Fig. 7. The No. of Utterances when Subjects Realized OOV

4. In-car noise speech evaluation

4.1 Speech data collected in driving car

For the robustness of ASR interface, we evaluate in-car noisy speech and propose an effective pre-processing technique to cope with in-car noise. We used in-car speech data which were collected in moving cars using array microphones (Waseda 2007). The recording was done in downtown Tokyo where the car was forced to drive slowly with frequent stops due to the traffic jam. Therefore, a large part of the background noise was from the surrounding environment such as other cars, constructions, etc. The speaker was sitting on the passenger seat, and there was a linear microphone array on the dashboard in front of the speaker shown Figure 8. The array consists of 7 (seven) microphones, which are located at the interval of 10cm, 5cm, 5cm, 5cm, 5cm, and 10cm. The array microphones were labeled as #1 to #7 from the driver seat side to the window side, so the central microphone was #4. Also, the headset microphone (#8) was used to collect noise-free speech.



Fig. 8. A Microphone Array in Car

4.2 Original data analysis

We have collected the speech data from 18 speakers (11 males and 7 females, all in their early twenties). 3,620 utterances for 152 POIs (Point of Interests) were collected in total, and they were roughly segmented using a fixed time period from the beep. After segmentation, the length of the data was approximately 7 hours in total. The number of utterances per speaker ranged from 134 to 326, and the number of utterances per POI ranged from 10 to 48. We then estimated the signal-to-noise ratio (SNR) by comparing the power of speech and non-speech segments. Table 4 shows the estimated SNR for each microphone. Since the noise spectrum has a strong peak in the low-frequency range, we also calculated the SNR after applying a bandpass filter of 400Hz to 5500Hz range. It is interesting that the estimated SNR does not have any correlations with distance between speakers and microphones, although speech recognition rate has correlations with distance of speakers and microphones. The recognition rate of headset microphone #8 is the best and would be the target rate. Figure 9 shows ASR recognition rate for each microphone after low power cut processing. The low power cut frequencies are 20Hz for males and 40Hz for females.

mic. ID	SNR (full band) dB	SNR (400-5500Hz) dB	Rec. rate(%)
1	-0.5	9.3	83.2
2	-2.8	12.1	86.3
3	-3.4	8.6	86.8
4	-3.0	9.2	87.8
5	-2.7	11.7	88.7
6	-3.8	8.5	85.6
7	-2.9	10.5	76.4
close-talk	56.7	83.2	95.2

Table 4. Estimated SNR of Each Microphone Data (Obuchi & Hataoka, 2006)

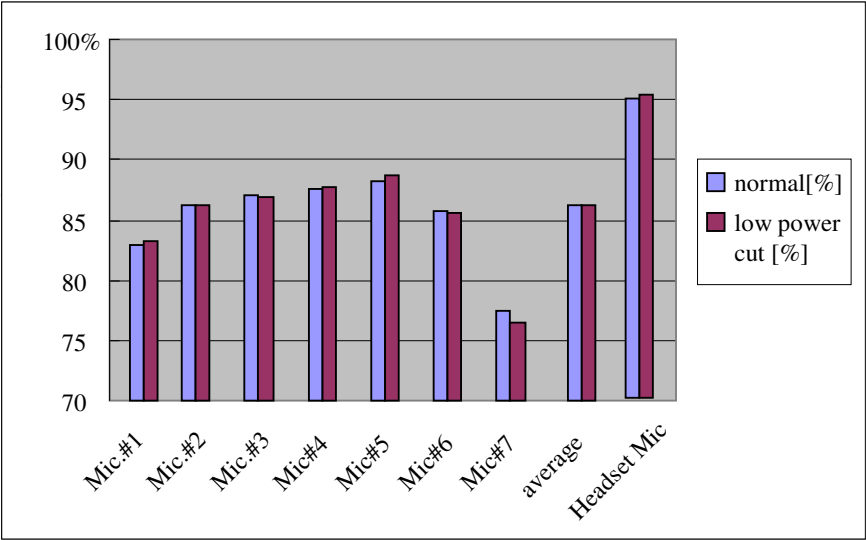


Fig. 9. Feature of Each Microphone

4.3 Experimental procedure

4.3.1 Free/open CSR software Julius/Julian

Free CSR (Continuous Speech Recognition) software Julian (Julius/Julian URL) was used as an ASR engine. There are two types of CRS engines such as Julian and Julius. The Julius is using language models based on N-gram and the Julian is using language models based on network grammars. The only difference between Julius and Julian is the language model. Both engines are using the same speech analysis and the same search algorithm. The search algorithm is based on two pass algorithm; the first search is a rough search using mono-phone HMMs and the second search by tri-phone HMMs. Julius is using bigram for the first search and trigram for the second search. We evaluate many noise handling techniques using a Julian decoder in automotive environments to make technical problems of speech processing modules clear.

4.3.2 Recognition experiments using various techniques

We carried out evaluation experiments of 152 POI isolated word recognition using Julian decoder for 5-male and 5-female data on a Linux machine. For Julian conditions, the sample acoustic model with PTM triphones was used. Among various variations of Julian, the Julian-v3.4.2 grammar driven decoder with 12 MFCC and log power, plus their first-order time derivatives is used. All the data were originally sampled by 44.1kHz, but down-sampled to 16kHz prior to the experiments.

The results of the baseline experiments (Table 1) showed the recognition rate of distant-talk was 88.7% (mic. #5) and that of close-talk was 95.2%. In the experiments, the individual recognition rate ranged from about 81% to 92% (average of all microphones).

The following pre-evaluation experiments are carried out to check problems of noisy speech data.

1. Evaluation of Low Power Cut and Spectrum Subtraction (SS)

According to many reports, the engine noise is ranging under 100Hz, so we carried out low power cut before the recognition stage. We checked frequency ranges of low power cut from 20Hz to 100Hz by 20Hz step size. Figure 10 shows results of low power cut (mic.#5). The cut of frequency 20Hz showed the best results (all average 88.8%), especially to female data.

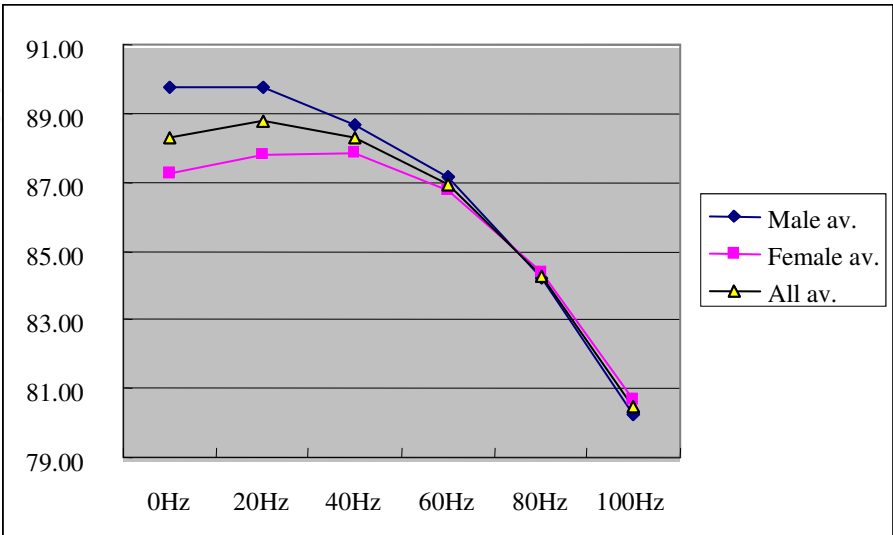


Fig. 10. Evaluation Experiments of Low Power Cut

There are many parameters for setting SS processing. We used standard function supported by the Julius/Julian software. Figure 11 shows all evaluation results, e.g. original data, low power cut data, and SS data (for all distant-talk microphones). These results show that the recognition rate and the pre-processing effects depend on talkers deeply.

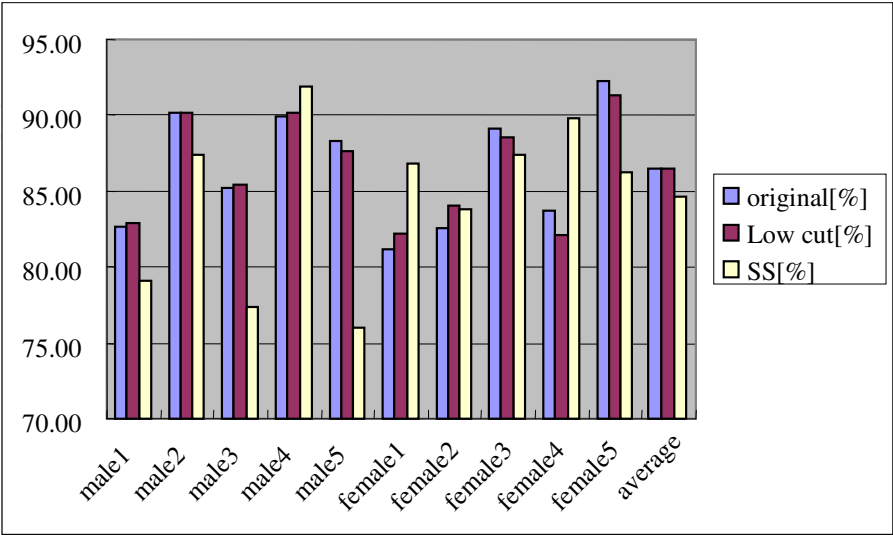


Fig. 11. Summary of Evaluation Experiments (Original/Low Power Cut/SS)

2. Weighted Summation of Array Microphones (WS) + Spectrum Subtraction (SS)

We tried the combination of array microphones and Spectrum Subtraction (SS). First, we summed speech data of possible array microphones, and then the SS processing was carried out to summed speech data. This array microphone technique is called Weighted Summation of Array Microphones (WS).

For SS, the following equation is used and two parameters  $a$ ,  $\beta$  are checked from the viewpoint of recognition accuracy.

$$S(f) = X(f) - \alpha \hat{N}(f)$$
$$\hat{N}(f) = (1 - \beta) \hat{N}'(f) + \beta N(f)$$

(1)

$\hat{N}(f)$ : estimated noise,  $\hat{N}'(f)$ : previous estimated noise  
 $\alpha$ : alpha parameter,  $\beta$ : floor parameter

In the Julian software, the default of  $a = 2.0$ , and the default of  $\beta = 0.5$ . We checked  $a = 2.0, 3.0, 3.5, 4.0$  and  $5.0$ .

For WS, three types of summation of microphones are used; microphones #3 + #4, microphones #4 + #5, and microphones #3 + #4 + #5.

Figure 12 shows recognition rate using the SS pre-processing method according to the value of parameter  $a$ . The pair of microphone #4 and #5, and the parameter  $a = 3.5$  gave the best recognition accuracy. For all pairs of #3, #4, and #5, the high average was obtained by  $a=4.0$ . However, the case of  $a = 3.5$  gave a recognition dip for the pair of #3 and #4.

Figure 13 shows recognition results according to pre-processing, i.e. SS, WS, and SS+WS. The headset microphone gave the best recognition accuracy (around 98%), and WS of microphone #4 and #5 (+#3), and SS gave the second best accuracy (around 90%). However, there is still a big gap between the headset and array microphones + SS pre-processing.

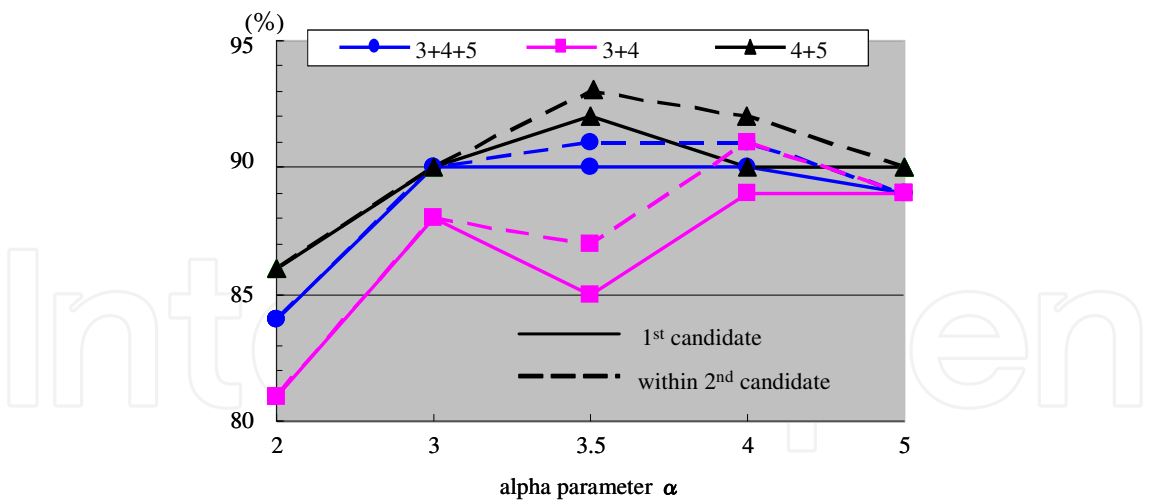


Fig. 12. Recognition Rate according to Alpha Parameter α

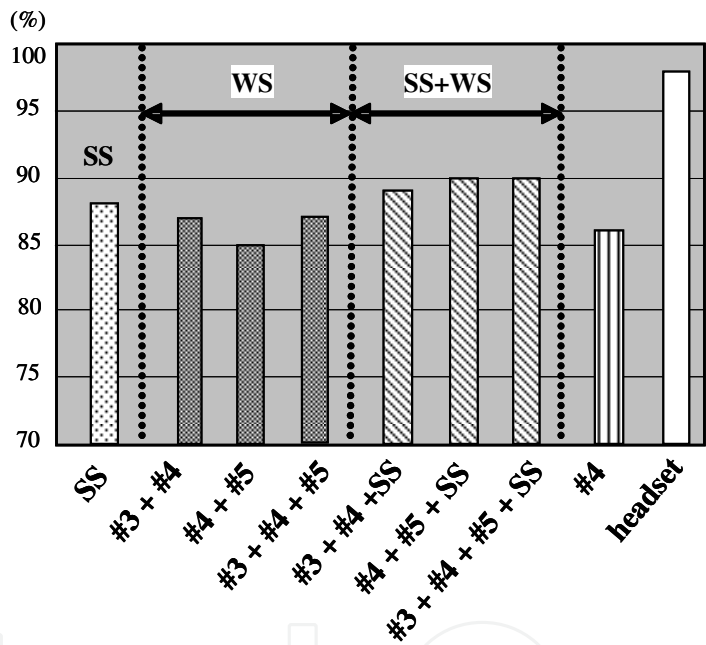


Fig. 13. Recognition Results according to Pre-processing

#### 4.3.3 Consideration

In this work, we carried out possible techniques of the signal processing level to get a robust noise reduction method. Especially, we tried Weighted Summation of Array Microphones and Spectrum Subtraction. We obtained improvement using WS + SS, however there is still a big gap of recognition rate between the headset (98%) and WS + SS (90%), i.e. 8%. The recognition rate by the headset is the target one, so the further analysis is needed to reach the target accuracy by the pre-processing techniques.

#### 5. Future work

The improvement of ASR interfaces and how to deal with OOV utterances are the most important and urgent issues to be solved. Also, more compact and more noise robust



embedded version of Julius should be developed. For the noise robust technique, we are trying the subtractive array-microphone method for the adaptive noise estimation. Moreover the use of Missing Feature Theory (Raj et. al., 2005) is a promising one for the noisy speech.

## 6. Conclusions

This chapter described two experimental evaluation results of ASR interface for mobile use, especially for car navigation applications. First, we evaluated the usability aspects of speech interface on a car navigation system and second, we evaluated in-car noisy speech by the various pre-processing techniques. For the first evaluation, we used a prototype which has a promising speech interface called FlexibleShortcuts and Select&Voice produced by Waseda University. To check OOV (Out-Of-Vocabulary) problems, we used the special case in that many location names have been changed by the town/city merging. This means that previous old location names are OOVs. For the location setting applications on the car navigation system, we found many undesirable OOV utterances occurred which made the interface worse. From the second experiment to check car-noise problems, we propose the combination method of Weighted Summation of Array Microphone (WS) + Spectrum Subtraction (SS) to increase recognition accuracy.

## 7. Acknowledgement

This research work was supported by the fund of “Fundamental Technology Development on Speech Recognition” organized by Japanese Ministry of Economy, Trade and Industry (METI). The speech data collected in the driving cars were obtained from Central Research Laboratory, Hitachi Ltd. by the research project supported by New Energy and Industrial Technology Development Organization (NEDO) in Japan.

## 8. References

- Kokubo, H, et al. (Oct. 2006). Embedded Julius: Continuous Speech Recognition Software for Microprocessor, MMSP2006, Canada
- Nakano, T, Fujii, S, and Kobayashi, K. (Dec. 2007). Extensible Speech Recognition System using Proxy-Agent, Proc. of ASRU2007, pp. 601- 606
- Hataoka, H, et al. (Jan. 2004). Robust Speech Dialog Interface for Car Telematics Service, Proc of IEEE CCNC2004, Las Vegas
- Vertanen, K. (Apr. 2008). Combining Open Vocabulary Recognition and Word Confusion Networks, ICASSP2008, SPE-P4, G7, Las Vegas
- Obuchi, Y. and Hataoka, N. (Sept. 2006). Development and Evaluation of Speech Database in Automotive Environments for Practical Speech Recognition Systems, Proc. of Interspeech2006, Pittsburgh, PA, USA
- Waseda University IT Institute. (2007). Advanced Research on Speech Recognition Technologies, C-3 to C-52
- Julius/Julian URL – an Open Source Large Vocabulary CSR Engine:-  
<http://julius.sourceforge.jp/>
- Raj, B. and Stern, R. M. (Sept. 2005). Missing-Feature Approaches in Speech Recognition, IEEE Signal Processing Magazine, Vol.22, No. 5, pp.101-116



## **Speech and Language Technologies**

Edited by Prof. Ivo Ipsic

ISBN 978-953-307-322-4

Hard cover, 344 pages

**Publisher** InTech

**Published online** 21, June, 2011

**Published in print edition** June, 2011

This book addresses state-of-the-art systems and achievements in various topics in the research field of speech and language technologies. Book chapters are organized in different sections covering diverse problems, which have to be solved in speech recognition and language understanding systems. In the first section machine translation systems based on large parallel corpora using rule-based and statistical-based translation methods are presented. The third chapter presents work on real time two way speech-to-speech translation systems. In the second section two papers explore the use of speech technologies in language learning. The third section presents a work on language modeling used for speech recognition. The chapters in section Text-to-speech systems and emotional speech describe corpus-based speech synthesis and highlight the importance of speech prosody in speech recognition. In the fifth section the problem of speaker diarization is addressed. The last section presents various topics in speech technology applications like audio-visual speech recognition and lip reading systems.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yasunari Obuchi, Teppei Nakano, Tetsunori Kobayashi and Nobuo Hataoka (2011). Speech Interface Evaluation on Car Navigation System - Many Undesirable Utterances and Sever Noisy Speech -, Speech and Language Technologies, Prof. Ivo Ipsic (Ed.), ISBN: 978-953-307-322-4, InTech, Available from: <http://www.intechopen.com/books/speech-and-language-technologies/speech-interface-evaluation-on-car-navigation-system-many-undesirable-utterances-and-sever-noisy-spe>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen